PB86-103082

Inferring Willingness to Pay for Housing Amenities from Residential Property Values

Iowa Univ., Iowa City

Prepared for

Environmental Protection Agency
Research Triangle Park, NC

Sep 85

U.S. Department of Commerce
National Technical Information Service
NTIS.

PB86-103082

EPA-600/7-85-034
September 1985

INFERRING WILLINGNESS TO PAY FOR HOUSING AMENITIES FROM RESIDENTIAL PROPERTY VALUES

by

Joel L. Horowitz
Departments of Geography and Economics
The University of Iowa
Iowa City, IA 52242

EPA Cooperative Agreement CR-810753-01-0

EPA Project Officer: J. David Mobley
Co-Project Officer: Ralph Luken

U.S. ENVIRONMENTAL PROTECTION AGENCY
Air and Energy Engineering Research Laboratory
Research Triangle Park, North Carolina 27711

and

Benefits Staff
Washington, D.C. 20460

AIR AND ENERGY ENGINEERING RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA (Please read Instructions on the reverse before completing)		
1. REPORT NO. EPA-600/7-85-034	2.	3. RECIPIENT'S ACCESSION NO. PB86 103082/AS
4. TITLE AND SUBTITLE Inferring Willingness to Pay for Housing Amenities from Residential Property Values		5. REPORT DATE September 1985 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Joel L. Horowitz		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Departments of Geography and Economics The University of Iowa Iowa City, Iowa 52242		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. CR810753-01
12. SPONSORING AGENCY NAME AND ADDRESS EPA, Office of Research and Development Air and Energy Engineering Research Laboratory Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED Final; 7/83 - 7/85 14. SPONSORING AGENCY CODE EPA/600/13

15. SUPPLEMENTARY NOTES AEERL project officer is J. David Mobley, Mail Drop 61, 919/541-2612. Co-project officer is Ralph A. Luken, PM220, EPA, Economic Analysis Div., Washington, DC 20460.

16. ABSTRACT The report describes a new model of consumers' choices in housing markets that incorporates three heretofore overlooked market considerations. The new model has been tested econometrically against the standard model, and the standard model has been strongly rejected. The new model gives estimates of buyers' willingness to pay for housing amenities that are significantly different from the estimates produced by the standard model. The standard hedonic model of consumers' choices in housing markets, which forms the basis of much environmental benefits analysis, fails to take into account three important characteristics of the bidding process through which houses are sold: (1) the prices of houses are established through sequential bidding, so sellers do not necessarily accept the highest bids that might be made for their houses; (2) ignorance of the market may lead potential buyers to make bids that are either higher or lower than those they would make if they had complete information; and (3) the distribution of bids for a house is truncated at the seller's asking price. These characteristics of housing markets imply that buyers cannot choose houses so as to maximize a deterministic utility function subject to a deterministic budget restraint as required by the standard model.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Pollution Residential Buildings Mathematical Models Estimating Value	Pollution Control Stationary Sources Property Values	13B 13M 12A 14G 05C
18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS (This Report) Unclassified 20. SECURITY CLASS (This page) Unclassified	21. NO. OF PAGES 74 22. PRICE

EPA Form 2220-1 (9-73)

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

The standard hedonic model of consumers' choices in housing markets, which forms the basis of much environmental benefits analysis, fails to take account of three important characteristics of the bidding process through which houses are sold. First, the prices of houses are established through sequential bidding, so sellers do not necessarily accept the highest bids that might be made for their houses. Second, ignorance of the market may lead potential buyers to make bids that are either higher or lower than the ones they would make if they had complete information. Finally, the distribution of bids for a house is truncated at the seller's asking price. These characteristics of housing markets imply that buyers cannot choose houses so as to maximize a deterministic utility function subject to a deterministic budget constraint as required by the standard model. This report describes a new model of consumers' choices in housing markets that incorporates the foregoing market characteristics. The new model has been tested econometrically against the standard model, and the standard model has been strongly rejected. The new model gives estimates of buyers' willingness to pay for housing amenities that are significantly different from the estimates produced by the standard model.

ACKNOWLEDGMENTS

I thank Allen Goodman for providing the data used in this work and for his comments on an earlier draft of this report, John Kennan for providing a computer program that implements the algorithm of Berndt, et al., and Scott Wisner for assisting with the computations.

FIGURES

Number		Page
1.1	Comparison of observed sale prices and sale prices predicted by the new model	7
1.2	Comparison of observed sale prices and sale prices predicted by the standard model	8
3.1	Comparison of observed sale prices with sale prices predicted by the new model	39
3.2	Comparison of observed sale prices with sale prices predicted by the standard model	40

TABLES

Number		Page
1.1	Estimates of buyers' willingness to pay for changes in houses' attributes	10
3.1	Estimation results for the new and standard models	37
4.1	Parameters of the hedonic price function	50
4.2	Parameters of the willingness to pay function	52

CONTENTS

Abstract iii

Acknowledgments vii

Chapter 1: Summary and Conclusions 1

Chapter 2: Formulation of the Models 13

Chapter 3: Empirical Tests of the Models 30

Chapter 4: Estimation of Buyers' Willingness to Pay for Housing Amenities 44

References 57

Appendices

A. Proof That Pr(h sold) $\leq$ 1 in (2.28) 59

B. Contents of the Data Set 62

C. Quality Control Evaluation Report 65

CHAPTER 1

SUMMARY AND CONCLUSIONS

1.1 INTRODUCTION

Econometric estimation of households' willingness to pay for environmental and other residential amenities often is based on observations of the effects of environmental and other factors on residential property values. The standard estimation procedure, which was described originally by Rosen [18], is based on the hypothesis that each household occupies a house that maximizes its utility subject to a budget constraint. It can be shown that subject to regularity conditions, utility maximization occurs when the household's marginal willingness to pay for each housing attribute (e.g., house size, air quality at the residential location, etc.) equals the marginal price of that attribute.

In the standard procedure, willingness to pay estimation takes place in two steps. First, a hedonic price function describing the relation between the observed market prices of houses and attributes such as size and air quality is estimated. The first partial derivative of the hedonic price function with respect to an attribute level yields the marginal price function of this attribute. The numerical marginal prices pertaining to a particular household can be obtained by substituting the values of the attributes of that household's house into the estimated price function. The resulting marginal price of the attribute is then treated as an observation of that household's marginal willingness to pay for the attribute. In the second estimation stage, the relation between marginal willingness to pay and relevant explanatory variables is specified up to a finite set of constant parameters. The parameter values are inferred from the "observed" marginal willingness to

pay values (i.e., the estimated marginal prices) using suitable econometric methods.

This hedonic pricing approach has been used in a variety of studies of willingness to pay for environmental improvements [2, 11, 17]. However, the method has some serious difficulties. Three that are particularly relevant to the research described in this report are:

1. The assumptions underlying the standard hedonic model imply that houses are sold to the households that are willing to pay the highest prices for them (i.e., houses are sold to the highest bidders). However, it is unlikely that houses are sold to the highest bidders in real housing markets. In real markets, houses normally are sold through sequential bidding processes. Bids must be accepted or rejected as they are received, and a seller is not able to assemble all bids that might be made before deciding which one to accept. Therefore, the seller does not necessarily accept the highest bid that might be made. Substantial transactions costs tend to prevent the rectification of erroneous decisions regarding the rejection and acceptance of bids.

2. Potential buyers bid for houses with little or no knowledge of the bids made by others. Therefore, buyers do not know the price at which a house will be sold in advance of the sale, and buyers have incomplete information about the house price components of their budget constraints. This makes it impossible for buyers to choose houses by maximizing utility subject to deterministic budget constraints. Transactions costs tend to prevent buyers from rectifying any resulting errors in bidding or in choosing houses.

3. The sale price of a house rarely exceeds the seller's asking price. Thus, the asking price effectively truncates the distribution of bids. If ignorance of the market or other factors cause sellers to establish asking

prices that are below the prices buyers are willing to pay, then truncation of the distribution of bids at the asking prices will prevent identification of the buyers willing to make the highest bids. Therefore, houses will not necessarily be sold to these buyers. Moreover, a process in which buyers choose houses by maximizing utility subject to budget constraints will not assign unique buyers to houses.

This report describes a model of consumers' choices in housing markets that incorporates the market characteristics just discussed. The new model and the standard model have been estimated and tested econometrically to determine which provides a better description of the operation of these markets. The formulation and testing of the new model and the implications of this model for willingness-to-pay estimation are summarized in this chapter. The remaining chapters provide a more detailed discussion of these matters. Chapter 2 presents the detailed mathematical formulations of the new and standard models. Chapter 3 presents the econometric tests of the models, and Chapter 4 presents the willingness-to-pay analysis.

1.2 FORMULATION OF THE MODELS

In this research, the new and standard models have been formulated as bidding models. Bidding models are housing market models in which the dependent variable is the sale price of a house and is modeled as a function of attributes of the house and the buyer. It is assumed in these models that sales occur as results of a bidding process.

The standard model used in this research is the well-known bid rent model of housing markets. Roughly speaking, this model assumes that houses are sold to the buyers who bid the most for them, and a house's sale price is the amount the highest bidder is willing to pay for that house. Under the customary assumptions of hedonic analysis, the bid rent model is equivalent to

the model in which buyers choose houses to maximize utility subject to a budget constraint. Moreover, under the same assumptions, willingness-to-pay analysis based on the bid rent model is equivalent to analysis based on Rosen's utility maximization approach. Thus, the standard market model can be tested using either the bid rent or the utility maximization (Rosen) form. The bid rent form is more convenient in the research discussed here.

In the version of the bid rent model used here, the probability density that a house with observed attributes X has sale price p_s is

$$\text{Density}(p_s) = f[p_s|W(X)], \tag{1.1}$$

where W(X) denotes the buyer's willingness to pay for a house with observed attributes X, and f is the function giving the probability density that p_s is the highest bid made for the house.

In the new model, the probability density that a house with observed attributes X and asking price P_a has sale price p_s is

$$\text{Density}(p_s) = Q(p_s|P_a)g[p_s|W(X)]/P(\text{sale}) \tag{1.2}$$

if $p_s < P_a$, where g is the probability density that the buyer makes a bid of p_s, $Q(p_s|P_a)$ is the probability that the seller accepts a bid of size p_s, and P(sale) is the probability that the house is sold at any price. The probability that $p_s = P_a$ is

$$\text{Prob}(P_a) = \{1 - G[P_a|W(X)]\}/P(\text{sale}), \tag{1.3}$$

where G is the cumulative probability distribution function corresponding to g. The probability that p_s exceeds P_a is zero. P(sale) is given mathematically by

$$P(\text{sale}) = \int_{\infty}^{P_a} \text{Density}(p)dp + \{1 - G[P_a|W(X)]\}. \tag{1.4}$$

Equations (1.2)-(1.4) constitute a bidding model of consumers' choices in housing markets that is analogous to the standard model but that is based on different assumptions concerning the bidding process that operates in housing

markets. In particular, (1.2)-(1.4) incorporate the hypotheses that sellers process bids sequentially, buyers do not have certain knowledge of the bids made by other buyers, and the distribution of bids for a house is truncated at the sellers's asking price. (See chapter 2 for a detailed explanation of how these hypotheses are represented in the model.) These hypotheses are inconsistent with the standard model's assumption that houses always are sold to the buyers willing to pay the most for them. As a consequence, it is possible to distinguish empirically between the new model [i.e., the model of equations (1.2)-(1.4)] and the standard model and to test each against the other empirically to determine which best explains the available data.

1.3 DATA

The new and standard models were tested and compared empirically using data describing a sample of 1196 houses sold in Baltimore City and County, Maryland, during 1978. The data were assembled from records of the Central Maryland Multiple Listing Service, Baltimore City and County tax records, the U.S. Census Bureau, the Maryland Bureau of Air Quality, and the results of previous investigations of school quality and crime levels in the Baltimore area. The data include the sale price, asking price, and annual property tax liability of each sampled house as well as 49 attributes of the house and its neighborhood. Environmental quality is characterized in these data by 3 air pollution variables: annual geometric mean particulate concentration, annual average NO_2 concentration, and annual maximum hourly O_3 concentration.

1.4 RESULTS

The new and standard models were both estimated (i.e., fit to the data) using standard econometric methods. These methods make it possible to test the models against one another by comparing how well each model fits the

data. If the fit of one model is significantly worse than the fit of the other, then it is possible to conclude with a high degree of confidence that the worse fitting model is incorrect.

Figures 1.1 and 1.2 illustrate the results of this comparison process graphically. In Figure 1.1, the observed sale prices of 50 randomly selected houses are plotted against the predicted sale prices obtained from the new model. In Figure 1.2, the observed sale prices of the same 50 houses are plotted against the predicted sale prices obtained from the standard model. The predicted sale prices were computed as expected values conditional on attribute levels and, in the case of the new model, asking prices. It can be seen from Figure 1.1 that the new model fits the observations well. There is no indication of serious errors in the model. It is clear from Figure 1.2 that the standard model fits the observations much more poorly than does the new model.

The differences in fit shown in Figures 1.1 and 1.2 are reflected in the root-mean-square (RMS) errors of the two models' predictions of houses' sale prices. For the 1196 houses in the full data set, the new model has a RMS prediction error of $2,830, whereas the standard model has a RMS prediction error of $11,360.

These results suggest strongly that the standard model is erroneous. It is possible to perform a formal statistical comparison of the new and standard models. This comparison leads to the conclusion that if the standard model is correct, then the observed goodness-of-fit results are events whose probabilities of occurrence are less than 0.0001. Thus, it is virtually certain that the standard model is erroneous and that the new model provides a better explanation of the operation of the Baltimore housing market than does the standard model. Since the bid rent model is equivalent to the utility

FIGURE 1.1 -- COMPARISON OF OBSERVED SALE PRICES

AND SALE PRICES PREDICTED BY THE NEW MODEL

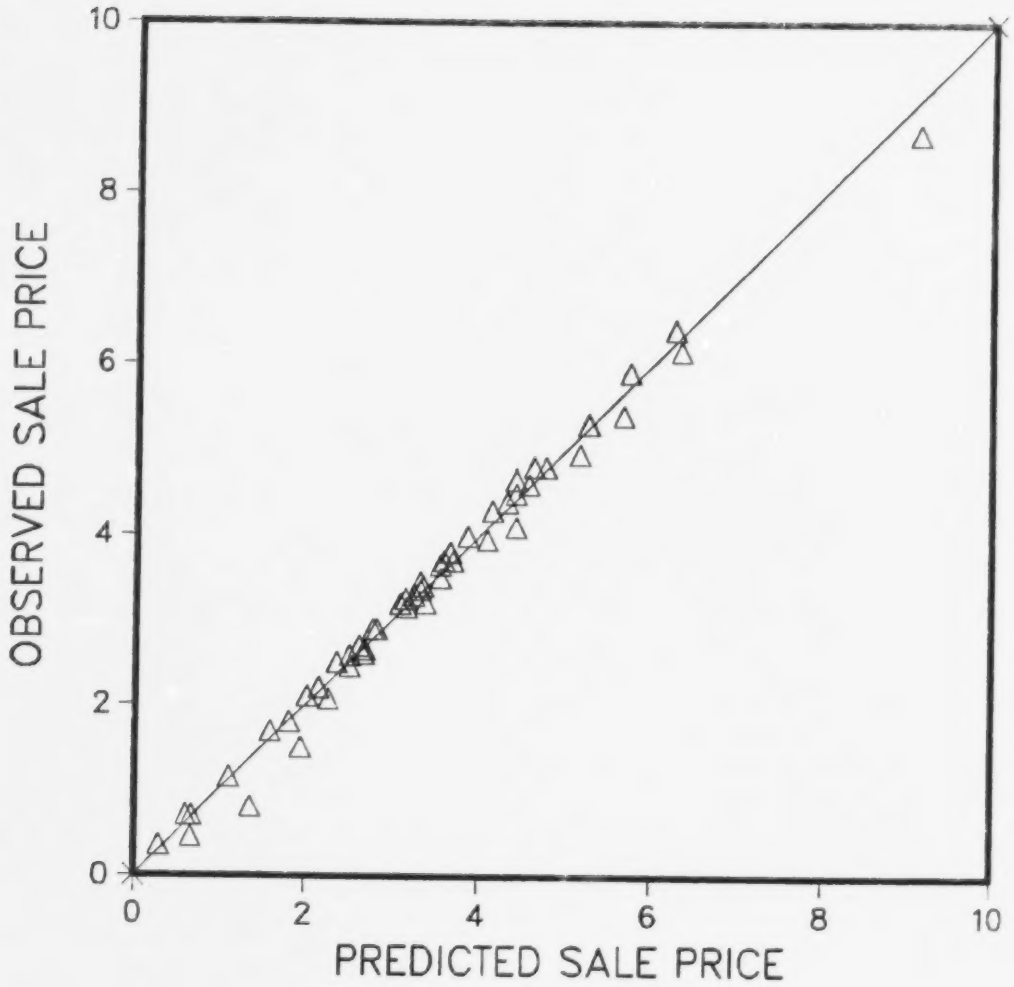
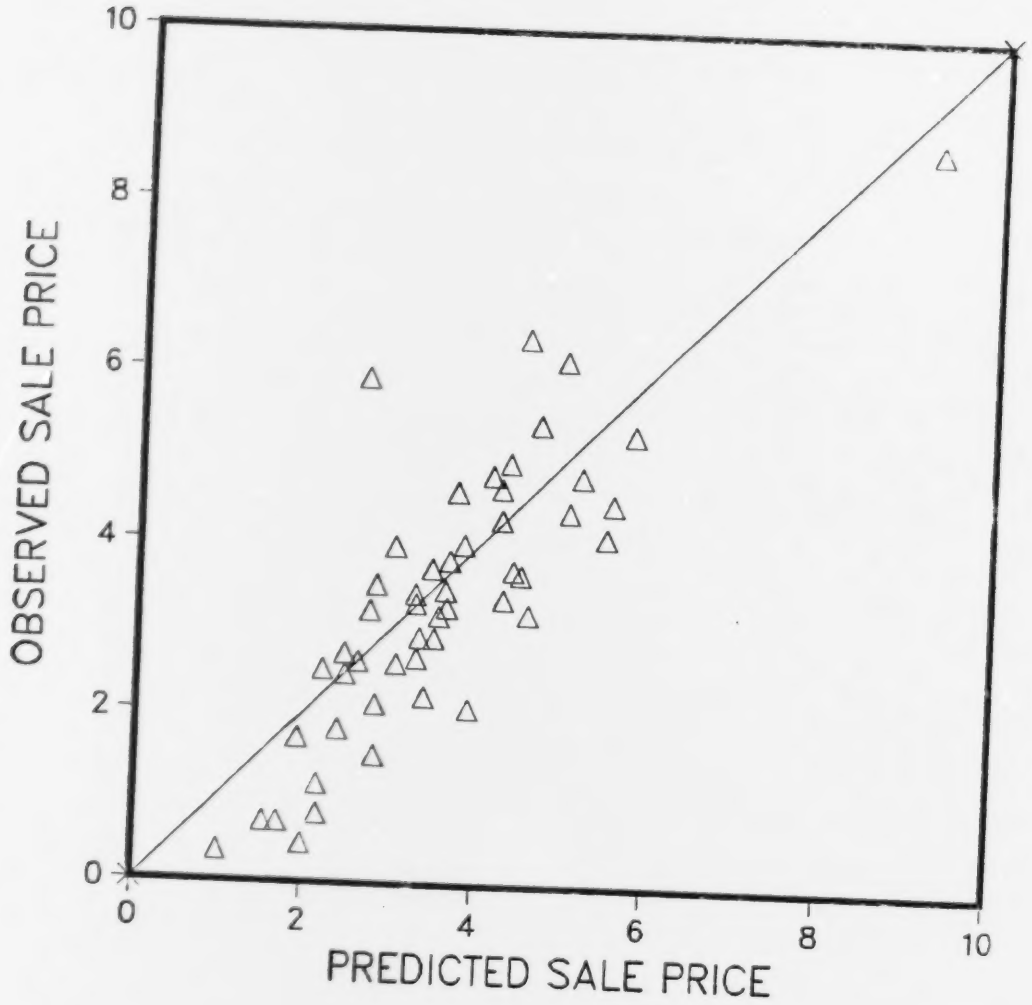


FIGURE 1.2 -- COMPARISON OF OBSERVED SALE PRICES AND SALE PRICES PREDICTED BY THE STANDARD MODEL



maximization model (in either the Rosen or random utility formulations, depending on whether housing is a continuous or discrete good), the latter model also is rejected.

The new and standard models give substantially different estimates of buyers' willingness to pay for changes in houses' attributes. Table 1.1 shows willingness to pay estimates for the following arbitrarily selected attribute changes: add a room, add 1000 $ft.^2$ to the lot, add a fireplace to a house that does not already have one, and add a garage to a house that does not already have one. The table also shows 95 percent confidence intervals for the willingness to pay estimates obtained from the new model. The contrast between the willingness to pay estimates obtained from the new and standard models is striking. Whereas the estimates obtained from the new model all are statistically significantly positive, as intuition suggests should be the case, all of the estimates obtained from the standard model are negative. Thus, the estimates obtained from the standard model are unreasonable as well as statistically distinct from the estimates obtained from the new model. The willingness-to-pay estimates illustrate the practical differences between the new and standard models and the practical consequences of attempting to estimate willingness to pay with the incorrect standard model.

The willingness to pay estimates given in Table 1.1 do not include willingness to pay for air quality improvement. This is because no statistically meaningful association was found between the air pollution variables and either the prices of houses in the Baltimore area or the willingness of buyers to pay for these houses. In estimates of the hedonic price function of houses in the Baltimore area, none of the air pollution variables was found to have a statistically significant association with sale price. Moreover, the coefficients of the particulate matter and O_3 variables have signs indicating, unreasonably, that more air pollution is preferred to

TABLE 1.1-- ESTIMATES OF BUYERS' WILLINGNESS TO PAY

FOR CHANGES IN HOUSES' ATTRIBUTES

	New Model		Standard Model
Attribute Change	Estimate	95 Percent Confidence Interval	Estimate
Add a room	$2,540	$1,960 - 3,130	$-470
Add 1000 ft.^{2} to lot	270	222 - 318	-72
Add fireplace	1,350	290 - 2,410	-1,150
Add garage	737	47 - 1,430	-423

less. Similarly, none of the air pollution variables has a statistically significant association with willingness to pay for houses as estimated from the new model, and the coefficient of the particulate variable has the incorrect sign. The coefficients of the air pollution variables all have correct signs in the willingness to pay function estimated from the standard model, and the association of the O_3 variable with willingness to pay is nominally statistically significant in this model. However, given the virtual certainty that the standard model is erroneous, little importance can be assigned to willingness to pay results obtained from this model. Thus, it is not possible on the basis of the research reported here to reach conclusions concerning the willingness of Baltimore area households to pay for improvements in air quality.

The causes of the poor performance of the air pollution variables in this research cannot be identified with certainty from the available data. However, one possible cause is error in the estimation of pollutant concentrations in the vicinities of houses. The concentration data reported by the Maryland Bureau of Air Quality pertain to concentrations at a set of fixed monitoring stations. The concentrations at individual houses had to be interpolated from the concentrations at the stations. Errors introduced in the interpolation could have caused the coefficients of the air pollution variables to be statistically nonsignificant and to have incorrect signs. Another possible cause of the poor performance of the pollution variables is that pollutant concentrations in the Baltimore area during 1978 were relatively low and varied little among geographical areas. Most monitoring stations reported annual geometric mean particulate concentrations below the particulate ambient air quality standard. Although several stations reported O_3 concentrations above the O_3 standard, the range of variation of O_3

concentrations in the available data is only $67\mu g/m^3$, which is too small to be perceptible to most people.

1.5 CONCLUSIONS

The standard hedonic model of consumers' choices in housing markets, which forms the basis of much environmental benefits analysis, assumes a market mechanism that is considerably different from the one that casual observation suggests operates in real housing markets. This report describes a new model that incorporates the main characteristics of the market mechanism that appears to operate in real housing markets. The new model has been tested against the standard model econometrically, and the standard model has been rejected strongly. Thus, the econometric results strongly support the hypothesis that the new model provides a better description of the operation of housing markets and, therefore, a firmer basis for environmental benefits analysis than does the standard model. The new and standard models give considerably different estimates of buyers' willingness to pay for changes in houses' attributes. Moreover, the standard model gives willingness-to-pay estimates that are highly unreasonable. Thus, the new and standard models are clearly distinguishable in terms of their fundamental assumptions, empirical performance, and practical consequences.

CHAPTER 2

FORMULATION OF THE MODELS

2.1 INTRODUCTION

The standard hedonic model of consumers' choices in housing markets, which forms the basis of much environmental benefits analysis, takes either of two forms. In one form, the utility maximization model (Muth [16]; Rosen [18]), households choose houses so as to maximize a utility function subject to a budget constraint. The utility function depends on attributes of houses and households and on the quantities of non-housing goods and services consumed. The budget constraint reflects the prices and ownership costs of houses, possibly expressed through a hedonic price function, and the prices of non-housing goods and services. In the other form, the bid rent model (Alonso [1]), each household offers a bid for each house that equals the household's willingness to pay for the house at a suitable reference utility level minus ownership costs. Houses are sold to the highest bidders. The two forms of the standard model are equivalent at equilibrium (Wheaton [19]), in which case a household's reference utility level in the bid rent model is the maximum utility the household can achieve at the equilibrium prices.

A large number of theoretical and empirical studies of housing markets have been based on the standard model. However, it is at least arguable that the assumptions about the operation of housing markets underlying this model are not satisfied in real markets, even at equilibrium. Houses normally are sold through sequential bidding processes. Bids must be accepted or rejected as they are received, and a seller is not able to assemble all bids that might be made before deciding which one to accept. Therefore, the seller does not necessarily accept the highest bid that might be made. Substantial transactions costs tend to prevent the rectification of erroneous decisions

regarding the rejection and acceptance of bids.[1] In addition, potential buyers bid for houses with little or no knowledge of the bids made by others. This ignorance of the market may cause buyers to make bids that are either higher or lower than those they would make with more complete information (i.e., buyers may underestimate or overestimate the maximum utility levels they can achieve in the market). In terms of the utility maximization form of the standard model, buyers have incomplete information about the house price components of their budget constraints. Transactions costs tend to prevent buyers from rectifying any resulting errors in bidding or in choosing houses. Finally, the sale price of a house rarely exceeds the seller's asking price. Thus, the asking price effectively truncates the distribution of bids. If ignorance of the market or other factors cause sellers to establish asking prices that are below the prices buyers are willing to pay, then truncation of the distribution of bids at the asking prices will prevent identification of the buyers willing to make the highest bids. Therefore, houses will not necessarily be sold to these buyers. In terms of the utility maximization model, utility maximization subject to a budget constraint will fail to assign unique buyers to houses.[2]

This chapter presents a bidding model of consumers' choices in housing markets that incorporates the market characteristics that have just been discussed. The chapter also presents the mathematical formulation of the standard model that will be used in the econometric tests of the two models described in Chapter 3.

2. THE STANDARD MODEL

The standard bid rent model is presented here in the form of a random bidding model (Ellickson [8]; Lerman and Kern [14]), since this form is convenient for the subsequent analysis. A random bidding model is an

econometrically estimable version of the bid rent model that uses a particular treatment of the stochastic component of the bid function. The nature of this treatment will become clear in the following discussion. Also for convenience, the model is formulated here in terms of actual sale prices of houses, rather than the annualized rents often used in housing models.

Let X denote a vector of attributes of houses (including neighborhood and locational attributes), Z a composite commodity with unit price that encompasses the present value of non-housing goods and services consumed throughout a household's planning period plus the present value of the household's financial assets at the end of the period, S a vector of attributes of households other than income and financial assets, Y a household's purchasing power (or wealth), $U(X,Z,S)$ the utility function of households, and $U_o(Y,S)$ the reference utility level of a household with attributes Y and S. Let U be a continuous, strictly increasing function of Z. Note that if R denotes the present value of a household's net expenditures for housing (including taxes; transaction, financing, and maintenance costs; and the difference between the purchase price of a house and its discounted resale value), then $Z = Y - R$.

A household's willingness to pay for a house with attributes X at reference utility level U_o, $w_o(X,Y,S)$, is defined as the solution to

$$U_o(Y,S) = U[X,Y - w_o(X,Y,S),S] \tag{2.1}$$

or, equivalently, by

$$w_o(X,Y,S) = Y - U^{-1}[X,S,U_o(Y,S)], \tag{2.2}$$

where U^{-1} denotes the inverse of $U(X,Z,S)$ with respect to Z. Thus, a household with attributes Y and S and reference utility level U_o will offer to

pay $w_o(X,Y,S)$ for a house with attributes X. However, in bidding for a house, a household commits itself to paying a variety of transaction and financing costs, property taxes, and maintenance costs in addition to the acutal purchase price of the house. Moreover, the household will recover the resale value of the house when it is sold in the future. Let C denote the discounted present value of transaction and financing costs, property taxes, and maintenance costs minus the discounted resale value of the house. Then a bid b entails a commitment to pay b + C, so the bid satisfies $b = w_o(X,Y,S) - C$. In general, C depends on the magnitude of the bid (e.g., because financing costs and property taxes are likely to depend on sale prices). If this dependence is linear, as will be assumed throughout this paper, then the value of C corresponding to a bid of b is $C(b) = C_o + C_1 b$, where C_o and C_1 may depend on X, Y, and S but are independent of b. It follows tnat b and w_o are related by

$$b = [w_o(X,Y,S) - C_o]/(1 + C_1). \tag{2.3}$$

To develop an empirical model of housing choice based on these ideas it is necessary to take account of the likelihood that available data do not include observations on all of the relevant elements of X and S. This can be done by treating w_o as the sum of deterministic and random components -- i.e.,

$$w_o = W_o(X,Y,S) + \varepsilon, \tag{2.4}$$

where W_o is the deterministic component of w_o, ε is the random component, and X and S now are redefined to refer to the observed attributes of houses and households. It can be assumed without loss of generality that the mean of ε is zero. It will also be assumed that the ε values pertaining to different households are independent.

Let ε be a continuous random variable with cumulative distribution function F_ε and density function f_ε. Then the cumulative distribution of bids b for a house with observed attributes X by households with observed attributes Y and S is $F_\varepsilon[(1 + C_1)b + C_o - W_o(X,Y,S)]$, and the corresponding density function is $(1 + C_1)f_\varepsilon[(1 + C_1)b + C_o - W_o(X,Y,S)]$. Suppose there are N households in the housing market and that the n'th household ($n = 1,\ldots, N$) has observed attributes Y_n and S_n. Since it is assumed in the standard model that houses are sold to the highest bidders, the probability density $g_h(i,p_s)$ that a house h with observed attributes X_h, C_{oh}, and C_{1h} is sold to household i at price p_s is

$$g_h(i,p_s) = (1 + C_{1h})f_\varepsilon[(1 + C_{1h})p_s + C_{oh} - W_o(X_h,Y_i,S_i)]$$

$$\cdot \prod_{\substack{n=1 \\ n\neq i}}^{N} F_\varepsilon[(1 + C_{1h})p_s + C_{oh} - W_o(X_h,Y_n,S_n)]. \qquad (2.5)$$

Equation (2.6) is a slightly generalized version of the random bidding model of housing markets developed by Lerman and Kern [14]. The model can be made econometrically estimable by specifying the functions F_ε and W_o up to finite sets of constant parameters. The parameter values then can be estimated by the method of maximum likelihood or, if the expected value of p_s conditional on h and i is a tractable function of the Y's, S's, C's, and X, by regression.

3. THE NEW MODEL

A household that is considering whether to bid for a particular house typically knows a vector of attributes of the house (X); the seller's asking price (P_a); and the property taxes, transaction costs, and financing costs that must be paid if the house is purchased at a specified price. The

household also is likely to have estimates of the maintenance costs and resale value of the house, although these may be highly subjective. Let $C(p_s) = C_o + C_1 p_s$ denote the discounted present value of transaction costs, property taxes, and estimated maintenance costs minus the discounted estimated resale value of the house when the sale price is p_s. In addition, the household is likely to know the sale prices of houses that are similar to the one being considered and that have been sold recently. This information may be used to estimate the likely resale value of the house, among other uses. The household typically does not know either the minimum bid acceptable to the seller or the price for which the house ultimately will be sold. However, the household can be expected to have formed subjective estimates of the likelihoods that the seller will accept bids of various sizes. These estimates, which almost certainly depend on P_a and which may also depend on such factors as the sale prices of recently sold similar houses, will be represented in this discussion by a function H(b), which is the household's subjective estimate of the probability that a bid b will be accepted.[3] Since houses rarely have sale prices above their sellers' asking prices, it will be assumed here that $H(b) = 1$ when $b \geq P_a$. Thus, the household has no incentive to make a bid that exceeds P_a.[4] Finally, there is usually a substantial penalty for reneging on an accepted bid. Hence, the household bids on only one house at a time.

The household must decide which house to bid on and how much to bid. It is assumed here that the household solves this problem by choosing the house and bid that maximize the expected value of a Von Neumann-Morgenstern utility function U. Thus, the household maximizes

$$E(U) = H(b)U[X, Y - C(b) - b, S] + [1 - H(b)]\, U_c, \tag{2.6}$$

where U_c is the household's current utility level. Let $w_c(X,Y,S)$ denote the household's willingness to pay for a house with attributes X at the utility level U_c.[5] Then w_c satisfies

$$U_c = U[X,Y - w_c(X,Y,S),S], \tag{2.7}$$

and E(U) satisfies

$$E(U) = H(b)U[X,Y - C(b) - b,S] + [1 - H(b)]U[X,Y - w_c(X,Y,S),S]. \tag{2.8}$$

Although it is possible in principle to develop a model of housing choice using a general utility function U and a general subjective probability function H, the theoretical and empirical analyses both are simplified greatly if certain assumptions are made about the forms of these functions. It is convenient to assume that U has the form

$$U(X,Z,S) = V(X,S) + Z. \tag{2.9}$$

In other words, the utility function is additively separable into housing and non-housing components, and households are risk-neutral with respect to choices among risky prospects involving the composite commodity Z at fixed X levels. Although these obviously are restrictive assumptions, equation (2.9) approximates any well-behaved utility function over a sufficiently small range of X and Z values and, as will be discussed later in this report, appears to yield a useful model. Equations (2.8) and (2.9) imply after some algebra that

$$E(U) = U_c + H(b)\,[w_c(X,Y,S) - C(b) - b]. \tag{2.10}$$

The form of H that will be used in the subsequent discussion is

$$H(b) = \begin{cases} \exp\,[(b - P_a)/\alpha] & \text{if } b \le P_a \\ 1 & \text{otherwise,} \end{cases} \tag{2.11}$$

where α is a positive constant. This form of H has no formal justification, but it is tractable, appears to yield a useful model, and has properties that are intuitively plausible. Specifically, H increases toward 1 as b approaches P_a and, depending on the value of α, H may be very small when b is substantially less than P_a.

Let b_{ih}^* denote the value of b that maximizes E(U) for household i and house h. Then b_{ih}^* is the amount that household i will bid for house h if i decides to make an offer for this house, or i's potential bid for house h. It follows from (2.10), (2.11), and the assumed form of C(b) that

$$b_{ih}^* = \begin{cases} [W_c(X_h, Y_i, S_i) - C_{oh}]/(1 + C_{1h}) - \alpha & \text{if this quantity does not exceed } P_{ah} \\ P_{ah} \text{ otherwise,} \end{cases} \quad (2.12)$$

where the subscripts h and i denote variables that are evaluated for house h and household i, respectively. As in the discussion in Section 2.2, W_c is represented as the sum of a deterministic component, W_c, that depends only on observed attributes of houses and households and a random component, ε, that accounts for the effects of unobserved attributes. Thus

$$b_{ih}^* = \begin{cases} [W_c(X_h, Y_i, S_i) - C_{oh} + \varepsilon]/(1 + C_{1h}) - \alpha & \text{if this quantity does not exceed } P_{ah} \\ P_{ah} \text{ otherwise,} \end{cases} \quad (2.13)$$

where the vectors X_h and S_i now are understood to contain only observed attributes of house h and household i. It will be assumed that the ε values for different houses but the same household and for different households regardless of house are independently and identically distributed with cumulative distribution function F_ε and density function f_ε. Accordingly, the probability density g_b of potential bids b_{ih}^* by household i for house h is

$$g_b(b_{ih}^*) = (1 + C_{1h}) f_\varepsilon[(1 + C_{1h}) b_{ih}^* + C_{oh} + (1 + C_{1h})\alpha - W_{cih}], \quad (2.14a)$$

if $b < P_a$, and the probability $G_b(P_{ah})$ that the potential bid equals P_{ah} is

$$G_b(P_{ah}) = 1 - F_\varepsilon[(1 + C_{1h})P_{ah} + C_{oh} + (1 + C_{1h})\alpha - W_{cih}], \qquad (2.14b)$$

where W_{cih} denotes $W_c(X_h, Y_i, S_i)$.

Let U_{ih}^* denote the value of E(U) corresponding to household i, house h, and bid b_{ih}^*. Then household i makes an offer for house h (i.e., conveys a bid to the owner of h) if

$$\max_{k \neq h} U_{ik}^* < U_{ih}^*. \qquad (2.15)$$

The left-hand side of (2.15) is random, owing to its dependence on the random variables ε_k ($k = 1,2,\ldots; k \neq h$). It is straightforward to derive an expression for the probability distribution of $\max_{k \neq h} U_{ik}^*$. However, such an expression is difficult to use in empirical analyses because it depends explicitly on the values of the variables X, C_o, C_1, and P_a for all houses for which household i has a non-zero probability of bidding. In a large city there may be over 10,000 such houses.[6] The problems of acquiring and handling data on such a large number of houses can be avoided by treating the variables X_k, C_{ok}, C_{1k}, and P_{ak} ($k \neq h$) as independent realizations of a random vector (X, C_o, C_1, P_a) sampled from a hypothetical population of houses. The variables U_{ik}^* ($k \neq h$) then constitute a sequence of independently and identically distributed random variables. When the U_{ik}^* have any of a large class of distributions (Gumbel [10]), the distribution of $\max_{k \neq h} U_{ik}^*$ converges to the Type I extreme value distribution as the number of houses in the sequence becomes large. The convergence can be very rapid (Cramer [6], Galambos [9]). Accordingly, it is reasonable to assume that for any ξ

$$\Pr(\max_{k \neq h} U_{ik}^* < \xi) = \exp\left\{-\exp[-(\xi - U_{ic} - \mu)/n]\right\}, \qquad (2.16)$$

where μ and η are constant parameters whose values may depend on the observed attributes of household i, and U_{ic} is the value of U_c corresponding to household i.

It follows from (2.16) that the conditional probability that household i makes an offer for house h given the values of b_{ih}^* and ε_{ih} is

$$\Pr(\text{offer} \mid b_{ih}^*, \varepsilon_{ih}) = \exp\left[-\exp[-(U_{ih}^* - U_{ic} - \mu)/\eta]\right]. \tag{2.17}$$

However, (2.10), (2.11), and (2.13) imply that

$$U_{ih}^* - U_{ic} = \begin{cases} \alpha(1 + C_{1h})H(b_{ih}^*) & \text{if} \quad b_{ih}^* < P_{ah} \\ W_{cih} - C_h(P_{ah}) - P_{ah} + \varepsilon_{ih} & \text{if } b_{ih}^* = P_{ah}. \end{cases} \tag{2.18}$$

Therefore,

$$\Pr(\text{offer} \mid b_{ih}^*, \varepsilon_{ih}) = \begin{cases} \exp\left[-\exp\left[-\dfrac{\alpha(1 + C_{1h})H(b_{ih}^*) - \mu}{\eta}\right]\right] \\ \qquad \text{if } b_{ih}^* < P_{ah} \\ \exp\left[-\exp\left[-\dfrac{W_{cih} - C_h(P_{ah}) - P_{ah} + \varepsilon_{ih} - \mu}{\eta}\right]\right] \\ \qquad \text{otherwise.} \end{cases} \tag{2.19}$$

If $b < P_{ah}$, the probability density of bids b offered by household i for house h, $g_{oih}(b)$, is

$$g_{oih}(b) = \int g_b(b)\Pr(\text{offer} \mid b, \varepsilon)dF_\varepsilon(\varepsilon \mid b), \tag{2.20}$$

where $F_\varepsilon(\varepsilon \mid b)$ is the conditional cumulative distribution of ε given b. The probability that household i makes an offer of P_{ah} for house h is

$$G_{oih}(P_a) = \int G_b(P_{ah}) \Pr(\text{offer} \mid P_{ah}, \varepsilon) dF_\varepsilon(\varepsilon \mid P_{ah}). \tag{2.21}$$

Since the value of ε is a deterministic function of b if $b < P_{ah}$ [see equation (2.13)], (2.20) implies that

$$g_{oih}(b) = (1 + C_{1h}) f_\varepsilon[b + C_h(b) + (1 + C_{1h})\alpha - W_{cih}]$$

$$\cdot \exp\left[-\exp\left[-\frac{\alpha(1 + C_{1h})H(b) - \mu}{\eta} \right] \right] \tag{2.22}$$

if $b < P_{ah}$. If $b = P_{ah}$, then $\varepsilon_{ih} \geq P_{ah} + C_h(P_{ah}) + (1 + C_{1h})\alpha - W_{cih}$. It follows from (2.14b) that the probability of this event is $G_b(P_{ah})$. Therefore, $f_\varepsilon(\varepsilon \mid P_{ah})$, the probability density function corresponding to $F_\varepsilon(\varepsilon \mid P_{ah})$, is

$$f_\varepsilon(\varepsilon \mid P_{ah}) = \begin{cases} 0 & \text{if } \varepsilon < P_{ah} + C_h(P_{ah}) + (1 + C_{1h})\alpha - W_{cih} \\ f_\varepsilon(\varepsilon)/G_b(P_{ah}) & \text{otherwise.} \end{cases} \tag{2.23}$$

Combining (2.14b), (2.19), (2.21), and (2.23), the probability that household i makes an offer of P_{ah} for house h is

$$G_{oih}(P_{ah}) = \int_{\varepsilon_o}^{\infty} f_\varepsilon(\varepsilon) \exp\left[-\exp\left[-\frac{W_{cih} - C_h(P_{ah}) - P_{ah} + \varepsilon - \mu}{\eta} \right] \right] d\varepsilon, \tag{2.24}$$

where $\varepsilon_o = P_{ah} + C_h(P_{ah}) + (1 + C_{1h})\alpha - W_{cih}$.

House h is sold to household i if the seller of h accepts i's bid. Let Q(b) denote the probability that a seller accepts a bid b given the values of P_a and any other relevant variables, and let $Q_h(b)$ denote this function evaluated for house h. Since the seller processes bids sequentially, $Q_h(b)$ is not necessarily the probability that b is the highest bid the seller might ever receive. It is assumed here that $Q_h(P_{ah}) = 1$. In other words, a seller always accepts a bid equal to his asking price.[7] Given this definition of Q_h, the probability density that house h is sold to household i at a sale price p_s is

$$g_h(i,p_s) = (1 + C_{1h})\, Q_h(p_s) f_\varepsilon[p_s + C_h(p_s) + (1 + C_{1h})\alpha - W_{cih}] \cdot \exp\left[-\exp\left[-\frac{\alpha(1 + C_{1h})H(p_s) - \mu}{\eta}\right]\right] \tag{2.25}$$

if $p_s < P_{ah}$. The probability that house h is sold to household i at a sale price P_{ah} is

$$G_h(i,P_{ah}) = \int_{\varepsilon_o}^{\infty} f_\varepsilon(\varepsilon)\exp\left[-\exp\left[-\frac{W_{cih} - C_h(P_{ah}) - P_{ah} + \varepsilon - \mu}{\eta}\right]\right] d\varepsilon, \tag{2.26}$$

where ε_o is as defined in (2.24).

Equations (2.25) and (2.26) constitute a bidding model of consumers' choices in housing markets that is analogous to the standard model (2.5) but that is based on different assumptions concerning the bidding process that operates in housing markets. In particular, equations (2.25) and (2.26) incorporate the hypotheses that sellers process bids sequentially, buyers are uncertain as to the utility levels they can achieve in the market, and the distribution of bids for a house is truncated at the seller's asking price.

The standard model of (2.5) and the new model of (2.25)-(2.26) imply different assumptions about the clearance of housing markets. In (2.5), each household bids on each house and a seller always accepts the highest bid offered. Therefore, each house on the market is sold with certainty. Accordingly, in (2.5) the probability that house h is sold satisfies

$$\Pr(h \text{ sold}) = \sum_i \int_{-\infty}^{\infty} g_h(i,p)dp = 1 \tag{2.27}$$

for each h, where the sum is over all households in the housing market. In (2.25) and (2.26) there is no assurance either that offers will be made for all houses or that each seller will accept one of the offers he receives. Therefore, it is possible for houses to remain on the market unsold. In (2.25) and (2.26), the probability that house h is sold is

$$\Pr(h \text{ sold}) = \sum_i [\int_{-\infty}^{P_{ah}} g_h(i,p)dp + G_h(i,P_{ah})]. \tag{2.28}$$

It is shown in Appendix A that under plausible assumptions about sellers' behavior that insure that a seller does not accept several bids for the same house, Pr(h sold) in (2.28) satisfies $\Pr(h \text{ sold}) \leq 1$.

The possibility that houses can remain unsold in the static model of (2.25) and (2.26) does not imply that houses necessarily remain unsold forever. The model implicitly applies to a short time interval during which each household in the housing market makes at most one offer for a house. The duration of this interval is unlikely to exceed several days. A house that remains unsold according to (2.25) and (2.26) remains unsold at the end of the time interval being modeled but may be sold in later time intervals.

Since the standard model's assumption that all houses are sold with certainty presumably refers to the operation of housing markets over relatively long time periods, it is not appropriate to test model (2.5) against a model such as (2.25)-(2.26) that applies to very short time

periods. This problem can be avoided by basing the test on the distribution of houses' sale prices over a relatively long time period, such as a year, and by excluding unsold houses from the test population. Assuming that the parameters of the models being tested are stable over the selected time period, the distributions of sale prices in a population that excludes unsold houses can be calculated from the models as the conditional distributions of houses' sale prices, given that the houses are sold. Let $g_h(i,p_s \mid \text{sale})$ denote the conditional probability density that house h is sold to household i at price p_s, given that house h is sold. Then in the standard model (2.5)

$$g_h(i,p_s \mid \text{sale}) = g_h(i,p_s)/\Pr(\text{h sold}) = g_h(i, p_s), \tag{2.29}$$

where $g_h(i,p_s)$ is given by (2.5), since $\Pr(\text{h sold}) = 1$ in this model. In the new model of (2.25) and (2.26),

$$g_h(i,p_s \mid \text{sale}) = g_h(i,p_s)/\Pr(\text{h sold}) \tag{2.30}$$

$$= g_h(i,p_s)/\sum_i [\int_{-\infty}^{P_{ah}} g_h(i,p)dp + G_h(i,P_{ah})] \tag{2.31}$$

if $p_s < P_{ah}$, where $g_h(i,p_s)$ and $G_h(i,P_{ah})$ are given by (2.25) and (2.26), respectively. The conditional probability that house h is sold to household i at price P_{ah} given that house h is sold is

$$G_h(i,P_{ah} \mid \text{sale}) = G_h(i,P_{ah})/\Pr(\text{h sold}) \tag{2.32}$$

$$= G_h(i,P_{ah})/\sum_i [\int_{-\infty}^{P_{ah}} g_h(i,p)dp + G_h(i,P_{ah})]. \tag{2.33}$$

Equations (2.29), (2.31), and (2.33) are the forms of the standard and new models used in the empirical analysis. Econometric methods for estimating and testing the new model are discussed in Chapter 3.

FOOTNOTES FOR CHAPTER 2

1. Examples of transactions costs affecting housing markets are brokerage and settlement fees, loan initiation fees, moving costs (psychic as well as monetary), and penalty payments for breaking sales contracts.
2. There can be little doubt that truncation of the distribution of bids occurs. For example, in the data set used in the empirical part of this report, 34 percent of the houses had sale prices exactly equal to the seller's asking price, and only 4 percent had sale prices above the asking price. Among the houses whose sale prices exceeded their asking prices, the mean excess of the former prices over the latter was 2.8 percent. Of course, these statistics do not prove that asking prices are below the prices buyers would be willing to pay, but they do suggest that such a hypothesis is worthy of consideration.
3. Under some conditions, the household may have knowledge of past bids that have been made for a house and of changes in P_a that the seller may have made. This is the case, for example, if the household is engaged in bargaining for a house and at least one round of offers and counter-offers has occurred. Such knowledge undoubtedly affects the household's estimates of bid acceptance probabilities, and it is interesting to speculate on how it might be incorporated explicitly into the function H. However, records of unsuccessful bids and of changes in P_a that may have occurred prior to the sale of a house were not available for use in the work reported here. Therefore, since the main objective of this paper is to develop an empirically testable model, any knowledge of past bids and changes in P_a that a household may have are treated as random effects that are subsumed in the function H.

4. The assumption that $H(b) = 1$ when $b \geq P_a$ implies that houses' sale prices never exceed their sellers' asking prices. Moreover, the probability that a seller receives simultaneously two or more bids equal to his asking price must be negligibly small since, otherwise, some bids equal to the asking price would be rejected. Although these assumptions do not describe housing markets exactly, they seem reasonable approximations in view of the observation that houses' sale prices rarely exceed their sellers' asking prices.

5. In general U_c and, therefore, w_c depend on attributes of the household's current house. This dependence is incorporated into the present model by including such attributes in the vector S. In addition, U_c depends on the household's expectations concerning the lengths of time required to purchase houses with various attributes at various prices. For example, U_c is likely to be larger if the household expects to be able to purchase an acceptable house quickly and at a low price than if it has the opposite perceptions. The dependence of U_c on households' expectations about the market is not modeled here. See Ioannides [13] for discussion of a possible modeling approach.

6. A buyer who is assisted by a broker affiliated with the multiple listing service chooses among all of the listed houses. The houses the buyer selects for inspection can be considered the outputs of a choice process in which listed houses with low expected utility values are eliminated.

7. As explained in footnote 4, the assumption that $Q_h(P_{ah}) = 1$ seems a reasonable approximation to the observed operation of housing markets. The function Q presumably depends on the seller's expectations concerning the rates at which he will receive bids of various sizes and on his time rate of preference as well as on the size of the current bid. A slightly

more detailed discussion of the process by which sellers decide whether to accept or reject offers is given in Appendix A. However, the development of a detailed model of sellers' behavior is beyond the scope of this research.

CHAPTER 3

EMPIRICAL TESTS OF THE MODELS

3.1 INTRODUCTION

This chapter describes the empircal tests of the new and standard models of housing markets. The chapter is organized as follows. Section 3.2 describes the data used in the tests, Section 3.3 describes the econometric methods that were used, and Section 3.4 describes the results. Section 3.5 states the conclusions that can be drawn from the tests.

3.2 DATA

The empirical analysis was based on data describing a sample of 1196 houses sold in Baltimore City and County, Maryland, during 1978. The houses were sampled randomly from those sold through the Multiple Listing Service in the Baltimore area during the first week of each month of 1978. Data describing the houses and their neighborhoods were assembled from records of the Central Maryland Multiple Listing Service, Baltimore City and County tax records, the U.S. Census Bureau, and the results of previous investigations of school quality and crime levels in the Baltimore area (Dubin and Goodman [7]). The data include the sale price, asking price, and annual property tax liability of each sampled house and its neighborhood.[1] A complete listing of the data elements is given in Appendix B.

The available data do not include attributes of buyers. Therefore, in the econometric analysis attributes of neighborhoods serve as surrogates for attributes of buyers to the extent that these are correlated. This makes it impossible to identify separately the effects of attributes of neighborhoods and attributes of buyers in the models.

The data listed in Appendix B include 46 elements other than air pollution variables that might reasonably be included in the vector X of housing attributes that enters the willingness to pay functions W_o and W_c. The air pollution variables were not used in this part of the research. To minimize the computational difficulties involved in working with a large number of attributes, the econometric analysis described in this chapter was carried out using a vector X consisting of 10 principal components of the 46 element attribute set. The 10 components explain 72 percent of the variance of the attribute data. No attempt was made to assign behavioral meanings or interpretations to the principal components.

Annual property taxes were computed using a procedure enacted into Maryland law in 1978. According to this procedure, property tax rates are 3 percent of market value in Baltimore City and 1.5 percent in Baltimore County. If the pre-sale tax liability of a house sold in 1978 was below the value obtained by applying the appropriate tax rate to the sale price, then the house's liability increases from the pre-sale value to the value based on the sale price at the rate of 15 percent per year.

The discounted present value of future property taxes was computed using a pre-tax discount rate of 8.5 percent, a rate of increase in houses' market values of 7.5 percent per year, and a time horizon of 11 years. The discount rate is within the range of long-term bond interest rates in 1978, the rate of increase in houses' market values is equal to the average rate of increase

nationally during 1969-1978, and the time horizon equals the mean time between moves of homeowning households in the Baltimore area. Adjustment for the deductibility of property taxes from federal and state income taxes was carried out for each house using the marginal tax rate corresponding to the mean income of the census tract in which the house is located. The relation between gross income and taxable income was estimated from records of income tax payments in 1978. The results of the empirical analysis reported here are not sensitive to the values of the parameters used in the property tax computation and do not change significantly when reasonable alternative choices of the parameter values are used.

The available data do not include information on arrangements for financing houses' sales. Accordingly, financing costs were estimated by assuming a 20 percent down payment, 9.5 percent mortgage interest rate, 25-year amortization period, and no loan initiation points. The assumed interest rate and loan initiation points equal the ones that prevailed in the Baltimore area in 1978. Adjustment of interest payments for the effects of income taxes was carried out as described in connection with property taxes. The results of the empirical analysis are not sensitive to reasonable variations in the assumptions made about financing arrangements.

The available data also do not include information on transaction and maintenance costs and resale values. Therefore, these were assumed to be estimable functions of houses' observed attributes. The specifications of these functions are discussed in the next section.

3.3 FUNCTIONAL SPECIFICATION AND ESTIMATION PROCEDURES

Practical computational considerations make it desirable to adopt simple specifications for W_o, W_c, and the functions used to estimate transaction and maintenance costs and resale values. For purposes of empirical testing of the

new and standard models, each of these functions was assumed to be linear in the principal components of the attribute data. (The results of the tests are not sensitive to the detailed specifications of the functions. See Sections 3.4 and 4.3 for illustrations of this insensitivity.) Since this specification makes it impossible to identify separately the parameters of the willingness to pay, transaction and maintenance cost, and resale-value functions, these four functions were combined into a single linear function W* that represents the sum of willingness to pay and estimated resale value minus the sum of estimated transaction and maintenance costs. The specification of W* is

$$W^*(X) = \theta_o + \sum_{k=1}^{10} \theta_k X_k, \tag{3.1}$$

where X denotes the vector of principal components of the attribute data, X_k denotes the k'th principal component, and the θ's are constant parameters.

The discussion of subsection 3.2 implies that the sum of the discounted present values of a house's financing costs and property taxes is a piecewise linear function of the house's sale price. This function was approximated by the linear function corresponding to the segment containing the observed sale price. Denote this linear function by $T(p_s)$, where

$$T(p_s) = T_o + T_1 p_s. \tag{3.2}$$

The values of the parameters T_o and T_1 pertaining to each house are available from the computation of financing costs and property taxes and, therefore, are known. It now follows that in the standard model

$$W_o(X,Y,S) - C(b) = W^*(X) - T(b), \tag{3.3a}$$

and in the new model

$$W_c(X,Y,S) - C(b) = W^*(X) - T(b). \tag{3.3b}$$

Therefore, equation (2.5) becomes

$$g_h(i,p_s) = (1 + T_{1h}) f_\varepsilon[p_s + T_h(p_s) - W_h^*] \prod_{\substack{n=1 \\ n\neq i}}^{N} F_\varepsilon[p_s + T_h(p_s) - W_h^*], \tag{3.4}$$

and equations (2.25) and (2.26) become

$$g_h(i,p_s) = (1 + T_{1h}) Q_h(p_s) f_\varepsilon[p_s + T_h(p_s) + (1 + T_{1h})\alpha - W_h^*] \cdot \exp\left[-\exp\left[-\frac{\alpha(1 + T_{1h})H_h(p_s) - \mu}{\eta} \right] \right] \tag{3.5}$$

$$G_h(i,P_{ah}) = \int_{\varepsilon_o}^{\infty} f_\varepsilon(\varepsilon) \exp\left[-\exp\left[-\frac{W_h^* - T_h(P_{ah}) - P_{ah} + \varepsilon - \mu}{\eta} \right] \right] d\varepsilon, \tag{3.6}$$

where $\varepsilon_o = P_{ah} + T_h(P_{ah}) + (1 + T_{1h})\alpha - W_h^*$, and the subscripts h denote functions evaluated for house h.

The function Q was specified as

$$Q(b) = \begin{cases} \exp[(b - P_a)/\beta] \text{ if } b \leq P_a \\ 1 \text{ otherwise,} \end{cases} \tag{3.7}$$

where β is a constant parameter. The rationale for this specification of Q is essentially the same as that given earlier for H.[2] The random components ε of the willingness to pay functions w_o and w_c were assumed normally distributed with means of zero.

Since the available data do not include attributes of individuals, the standard model now takes the form

$$g_h(i,p_s) = [(1 + T_{1h})/\sigma]\phi\left[\frac{p_s + T_h(p_s) - W_h^*}{\sigma}\right]\left[\Phi\left[\frac{p_s + T_h(p_s) - W_h^*}{\sigma}\right]\right]^{N-1} \tag{3.8}$$

where σ is the standard deviation of ε, and Φ and ϕ, respectively, denote the cumulative standard normal distribution function and the standard normal density function. The right-hand side of (3.8) corresponds to the probability density of the maximum of N independently and identically normally distributed random variables. Therefore, since N is large (it is the number of households in the housing market), $g_h(i, p_s)$ can be approximated with very high accuracy by a type I extreme value density function:

$$g_h(i,p_s) = \omega(1 + T_{lh})\exp\left[-\omega[p_s + T_h(p_s) - W_h^* - m]\right] \cdot \exp\left[-\exp\left[-\omega[p_s + T_h(p_s) - W_h^* - m]\right]\right] \quad , \qquad (3.9)$$

where m and ω are constant parameters. Since (3.8) and (3.9) are equivalent when N is large but (3.9) is more tractable computationally than is (3.8), the standard model was estimated in the form (3.9). The parameter m is not identifiable and has been absorbed into θ_o.

The standard and new models were estimated by the maximum likelihood method using the algorithm of Berndt, et al. [3]. Preliminary estimation of the new model led to the conclusion that $\Pr(\text{offer} \mid b_{ih}^*, \varepsilon_{ih})$ is virtually constant in the estimation data set and, therefore, contributes little to the model's explanatory power.[3] Accordingly, $\Pr(\text{offer} \mid b_{ih}^*, \varepsilon_{ih})$ has been treated as a constant. This, together with the assumed normality of ε, causes (2.31) and (2.33) to take the form

$$g_h(i,p_s \mid \text{sale}) = \sigma^{-1}(1 + T_{lh})Q_h(p_s)\phi\left[\frac{p_s + T_h(p_s) + (1 + T_{lh})\alpha - W_h^*}{\sigma}\right] /K \qquad (3.10)$$

when $p_s < P_{ah}$, and

$$G_h(i,P_{ah} \mid \text{sale}) = \left\{1 - \Phi\left[\frac{P_{ah} + T_h(P_{ah}) + (1 + T_{lh})\alpha - W_h^*}{\sigma}\right]\right\}/K \qquad (3.11)$$

when $p_s = P_{ah}$, where

$$K = \int_{-\infty}^{P_{ah}} \sigma^{-1}(1 + T_{1h})Q_h(p)\phi\left[\frac{p + T_h(p) + (1 + T_{1h})\alpha - W_h^*}{\sigma}\right] dp$$

$$+ \left\{1 - \Phi\left[\frac{P_{ah} + T_h(P_{ah}) + (1 + T_{1h})\alpha - W_h^*}{\sigma}\right]\right\} . \qquad (3.12)$$

The form (3.10)-(3.12) is much more tractable computationally than are (2.31) and (2.33), and the estimation results reported here are based on (3.10)-(3.12).

3.4 ESTIMATION RESULTS

Table 3.1 shows the results of estimating the new and standard models. The most important entries in this table are the values of the two models' log likelihood functions.[4] The log likelihood of the new model exceeds that of the standard model by 1191.47. The significance of this difference can be understood as follows. Let $f(x,\psi)$ and $g(x,\theta)$ denote two hypotheses as to the distribution function of a random variable x, where ψ and θ are vectors of constant parameters, and assume that f and g are non-nested (i.e., that neither function can be obtained as a parametric special case of the other). Let L_f and L_g, respectively, denote the maximum values of the log likelihood functions of f and g, M denote the size of the estimation data set, and k_f and k_g denote the numbers of elements in the vectors ψ and θ. Suppose that model g is correctly specified and model f is misspecified. Then, as shown by Horowitz [12], asymptotically

$$\Pr\left[M^{-1/2}(L_f - L_g) \geq z\right] \leq \Phi\left[-(2M^{1/2}z - k_f + k_g)^{1/2}\right] , \qquad (3.13)$$

TABLE 3.1--ESTIMATION RESULTS FOR
THE NEW AND STANDARD MODELS

	New Model		Standard Model	
Parameter	Value	Standard Error of Estimate	Value	Standard Error of Estimate
θ_0	-7.43	6.85	0.146	0.214
θ_1	-0.181	0.245	-0.276	0.0628
θ_2	1.11	0.241	1.63	0.0422
θ_3	-1.02	0.168	-0.550	0.0438
θ_4	-0.0260	0.224	-0.138	0.0606
θ_5	0.387	0.217	0.363	0.0391
θ_6	-0.190	0.0955	-0.0761	0.123
θ_7	-0.249	0.228	0.102	0.0676
θ_8	1.09	0.307	0.151	0.0680
θ_9	0.422	0.237	0.270	0.0610
θ_{10}	0.514	0.245	0.112	0.0509
α	0.0520	5.33	NA	NA
β	0.189	0.00739	NA	NA
σ or ω	1.79	0.0773	0.838	0.0111
Log likelihood	-609.72		-1801.19	

Price units are tens of thousands of dollars.

where z is any positive number and Φ denotes the cumulative standard normal distribution function. Let M = 1196, k_f = 14, k_g = 12, and z = 0.2289. Then (3.13) becomes

$$\Pr(L_f - L_g \geq 7.92) \leq 0.0001. \tag{3.14}$$

Since the standard model estimated here contains 12 parameters and the new model contains 14, (3.14) applies to the difference between the two models' log likelihoods with g corresponding to the standard model. Thus, if the standard model is correct, there is virtually no chance that the log likelihood of the new model exceeds that of the standard model by more than 7.92. Hence, the observation that the new model's log likelihood exceeds the standard model's log likelihood by 1191.47 implies that the standard model is misspecified and that the new model provides a better explanation of the available data than does the standard model. Since the random bidding version of the standard model, used here, is equivalent to the Rosen or random utility versions, the latter versions also are rejected.

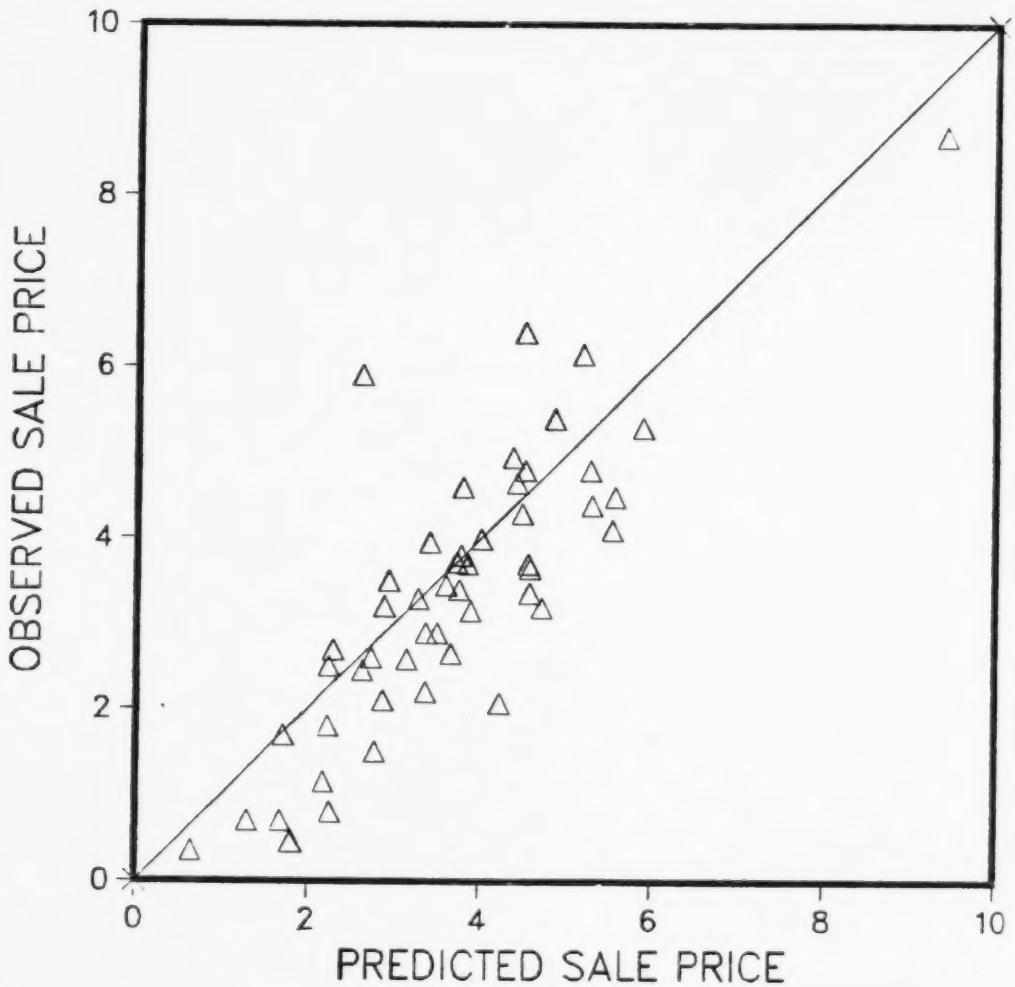
The new and standard models are compared graphically in Figures 3.1 and 3.2. In Figure 3.1, the observed sale prices of 50 randomly selected houses are plotted against the predicted sale prices obtained from the new model. In Figure 3.2, the observed sale prices of the same 50 houses are plotted against the predicted sale prices obtained from the standard model. (Figures 3.1 and 3.2 differ slightly from Figures 1.1 and 1.2. Figures 3.1 and 3.2 are based on equation (3.1), whereas Figures 1.1 and 1.2 are based on a specification that permits separate indentification of the willingness-to-pay function. See Chapter 4 for the details of the latter specification.) The predicted prices were computed as expected values conditional on attribute levels and, in the case of the new model, asking prices. It can be seen from Figure 3.1 that the new model fits the

FIGURE 3.1 -- COMPARISON OF OBSERVED SALE PRICES WITH SALE PRICES PREDICTED BY THE NEW MODEL

Based on a random sample of 50 houses. Prices are in tens of thousands of dollars. Diagonal line denotes locus of equal predicted and observed prices.

FIGURE 3.2 -- COMPARISON OF OBSERVED SALE PRICES WITH SALE PRICES PREDICTED BY THE STANDARD MODEL

Based on a random sample of 50 houses. Prices are in tens of thousands of dollars. Diagonal line denotes locus of equal predicted and observed prices.



error, despite the simple specifications that have been used in developing this model. It is clear from Figure 3.2 that the standard model fits the observations much more poorly than does the new model.

The differences in fit shown in Figures 3.1 and 3.2 are reflected in the root-mean-square (RMS) errors of the two models' predictions of houses' sale prices. For the houses in the full estimation data set, the new model has a RMS prediction error of $2,810, whereas the standard model has a RMS prediction error of $12,500.

The foregoing results suggest that the new model provides a substantially better representation of the bidding process operating in housing markets than does the standard model. However, two alternative explanations for the results also require consideration. First, it is possible that the asking price acts as a proxy for housing attributes not otherwise represented in the estimation data set and, thereby, causes the performance of the new model to be better than that of the standard model. To test this possibility, W* in the standard model was respecified as

$$W^* = \theta_o + \sum_{k=1}^{10} \theta_k X_k + \theta_{11} P_a. \qquad (3.15)$$

This has the effect of including in W* any housing attributes for which the asking price is a proxy. Re-estimation of the standard model with W* specified as in (3.15) yielded a log likelihood of -1035.81. Since the log likelihood of the new model exceeds this value by 426.10, it follows from inequality (3.13) that the standard model, even with the expanded specification of W*, is rejected in a test against the new model. The RMS prediction error of the standard model with the specification (3.15) is $4,320. Thus, the differences in the log likelihood values and RMS prediction errors of the new and standard models cannot be explained by the differences

in the models' explanatory variables but, rather, are due to the differences in their functional specifications.

The second alternative explanation of the poor performance of the standard model is that W* is seriously misspecified and that this misspecification causes the model to appear to perform poorly. This possibility was investigated by re-estimating the standard model using the following relatively flexible specification of W*:

$$W^* = (\theta_o + \sum_{k=1}^{10} X_k + \theta_{11} P_a)^{\lambda} , \qquad (3.16)$$

where λ is a parameter to be estimated. Equation (3.16) is equivalent to setting a Box-Cox transformation of W* equal to a linear function of X and P_a with P_a used as a proxy for housing attributes not represented in X. The resulting log likelihood is -1033.49, and the RMS prediction error is $4,320. Thus, the standard model continues to be rejected in a test against the new model.

3.5 CONCLUSIONS

This chapter has presented the results of empirical tests of the new hedonic model of housing markets, developed in Chapter 2, against the standard model. The results indicate strongly that the standard model is erroneous. It follows that analyses based on the standard model are likely to be misleading. The empirical results also indicate that the new model provides a considerably better description of the operation of housing markets than does the standard model. The practical implications of these findings are explored in Chapter 4.

FOOTNOTES FOR CHAPTER 3

1. The neighborhood of a house is defined variously as the census tract, school district, or local crime reporting area in which the house is located, depending on the data element in question.

2. This form of Q is adopted as a simple approximation to the more highly structured bid acceptance model described in Appendix A. The approximation consists of ignoring the possibility that several offers are made simultaneously for the same house. Although it is theoretically possible to estimate (2.31) and (2.33) using the model of Appendix A, practical comptutational considerations make the use of a simpler formulation desirable.

3. In other words, $H(p_s)(W^* - T - p_s)$ varies little among houses when p_s corresponds to actual sale prices.

4. There are no a priori expectations concerning the signs or significance levels of the coefficients of the principal components and no behavioral interpretations that can be assigned to the coefficients since the principal components have no behavioral interpretations and the coefficients pertain to the composite function W^*.

CHAPTER 4

ESTIMATION OF BUYERS' WILLINGNESS TO PAY FOR HOUSING AMENITIES

4.1 INTRODUCTION

The specifications of the new and standard models used in Chapter 3 are adequate for testing these models, but they do not permit buyers' willingness to pay for housing amenities to be estimated. This is because the willingness to pay function is not identified in these specifications. In this chapter, the specifications of the new and standard models will be modified so as to permit identification of the willingness to pay function. The modifications do not alter the results of the tests described in Chapter 3 or the interpretations of these results that were given in that chapter. Section 4.2 presents the modified specifications. Section 4.3 presents the results of estimating willingness to pay with the new and standard models. Section 4.4 presents the conclusions that can be drawn from this research.

4.2 SPECIFICATIONS OF THE MODELS

In the empirical analysis described in Chapter 3, the willingness-to-pay functions, W_o and W_c, and the functions used to estimate transactions costs, maintenance costs, and resale values all were assumed to be linear in the principal components of the attribute data. This assumption made it impossible to identify the willingness to pay functions. The objective of this section is to develop specifications for the willingness-to-pay, cost, and resale-value functions that do permit identification of willingness to pay. The specifications will include air pollution variables so as to make possible investigation of willingness to pay for air quality improvements.

Two problems must be dealt with in developing identifiable specifications. The first, which is the more serious of the two, concerns

whether the willingness-to-pay function is identifiable, even with an ideal data set. The second concerns the conditions under which the willingness-to-pay function is identifiable given the available data, which are not ideal for reasons that will be discussed presently.

The first identification problem arises from the observation of Brown and Rosen [5] that the the willingness-to-pay function in the standard hedonic model can be identified only if the specifications of the hedonic price and willingness-to-pay functions are restricted in appropriate ways. The identifying restrictions cannot be tested empirically using market data and, therefore, must be imposed arbitrarily. The cause of this identification problem is that if, as is assumed in the standard model, buyers choose houses so as to maximize utility subject to a budget constraint, then the observed values of houses' attributes and sale prices always correspond to points of tangency of the hedonic price and willingness-to-pay functions. Since the hedonic price function obviously is tangent to itself, separate identification of the hedonic price and willingness-to-pay functions is possible only if their specifications are restricted to suitably distinct parametric families of functions. No market data will make it possible to avoid these restrictions, and no market data will make the restrictions testable if the utility maximization paradigm underlying the standard model is correct. Note that the use of multi-market data does not solve this identification problem, since the willingness-to-pay function is identifiable with multi-market data only if the untestable assumption is made that this function is the same in all markets.

The new housing model does not assume that buyers choose houses so as to maximize utility subject to a budget constraint. Therefore, the tangency condition that causes lack of identification in the standard model does not

occur in the new model. The willingness-to-pay function in the new model is identifiable if appropriate data are available. Thus, the first identification problem does not arise in connection with the new model.

Unfortunately, the data available for this research (the Baltimore data) are not sufficient to permit identification of the willingness-to-pay functions of either the new or standard models unless restrictions are placed on the specifications of certain functions in these models. This is the second identification problem. It is present in both the new and standard models and is a consequence of the available data, rather than of the characteristics of the models. The second identification problem is conceptually distinct from the first problem, which affects only the standard model and is a consequence of the structure of that model, rather than of a particular data set.

The second identification problem is caused by the fact that the available data do not include observations of transactions costs, maintenance costs, and buyers' estimates of houses eventual resale values. Therefore, these quantities must be estimated as functions of houses' attributes. However, it is clear from equations (2.5), (2.25), and (2.26) that unless the willingness-to-pay function and the function estimating transactions costs, maintenance costs, and resale values belong to suitably distinct parametric families, the willingness-to-pay function cannot be identified. Accordingly, it is necessary to restrict the specifications of these functions to achieve identification of willingness to pay. It is easy to see that these restrictions would not be needed if the available data included observations of transactions costs, maintenance costs, and resale values, all of which are observable in principle. This is why the second identification problem is a consequence of the available data, rather than of the models being

considered. (It is worth noting that the second identification problem affects all empirical hedonic analyses. However, the problem is not visible in the results reported in the literature, where it is common to work in terms of imputed rents of houses without discussing the assumptions about costs and resale values needed to compute these rents.)

To achieve identification of the willingness-to-pay functions of both the new and standard models, it has been assumed for the purposes of this chapter that the sum of transactions and maintenance costs minus the discounted expected resale value of a house is proportional to the expected value of a hedonic price function. In accordance with current practice, the hedonic price function is specified as

$$p_s^{(\lambda)} = a_0 + \sum_{i=1}^{10} a_i X_i + \sum_{i=1}^{3} b_i V_i + \varepsilon, \tag{4.1}$$

where $p_s^{(\lambda)}$ denotes the Box-Cox transformed observed sale price, X_i denotes the i'th principal component of the attribute data excluding air pollution variables, V_i (i = 1,...,3) denotes the i'th air pollution variable (these variables will be explained in more detail presently), ε is a normally distributed random variable whose mean is zero, and the a's, b's, and λ are constant parameters whose values are to be estimated. The willingness-to-pay function was specified to be linear in the principal components and pollution variables:

$$W(X,V) = \sum_{i=1}^{10} c_i X_i + \sum_{1=1}^{3} d_i V_i, \tag{4.2}$$

where the c's and d's are constant parameters whose values are to be estimated. Of course, equations (4.1) and (4.2) are not the only possible specifications of the hedonic price and willingness-to-pay functions, although they are the most tractable specifications that permit identification of W.

The use of other specifications would lead to empirical results that differ in detail from the ones obtained with equations (4.1) and (4.2). However, the use of other specifications would not alter the main conclusions of this research, which are that the standard hedonic model is almost certainly erroneous, the new model provides a better description of the operation of housing markets than does the standard model, and the new and standard models yield substantially different estimates of buyers' willingness to pay for housing amenities.

The air pollution variables included in equations (4.1) and (4.2) are annual average particulate matter and NO_2 concentrations and annual maximum hourly O_3 concentration. Observations of these concentrations are available only at a set of fixed monitoring sites. There are 18 such sites for particulates, 11 for NO_2, and 4 for O_3. Pollutant concentrations at individual houses were estimated by interpolating among the observations at the fixed sites. The following interpolation formula was used:

$$V_{ih} = \sum_s w_{ihs} V_{is}, \tag{4.3}$$

where V_{ih} denotes the estimated concentration of pollutant i at house h, V_{is} denotes the concentration of pollutant i at monitoring site s, w_{ihs} is the interpolation weight for pollutant i, house h, and monitoring site s, and the sum extends over all monitoring sites at which pollutant i is observed. The weights are inversely proportional to the distances between monitoring sites and houses. Thus,

$$w_{ihs} = d_{hs}^{-1} / \sum_s d_{hs}^{-1}, \tag{4.4}$$

where d_{hs} is the Euclidean distance from house h to monitoring site s and the sum in the denominator extends over sites at which pollutant i is observed.

Estimation of the parameters of the hedonic price and willingness-to-pay functions was carried out by a two stage maximum likelihood procedure. In the first stage, the parameters of the hedonic price function were estimated by applying the maximum likelihood method to equation (4.1). In the second stage, the parameters of the willingness-to-pay function were estimated by applying the maximum likelihood to equation (3.9) for the standard model and equations (3.10)-(3.12) for the new model after replacing W^* in these equations by the sum of W in equation (4.2) and a constant multiple of the estimated expected value of the hedonic price function.[1] The constant multiplier, which is the proportionality constant relating the estimated value of the hedonic price function to the sum of transactions and maintenance costs minus resale value, was estimated as part of the maximum likelihood procedure.

4.3 RESULTS

Table 4.1 shows the estimation results for the hedonic price function. The estimated value of λ, 0.506, confirms the desirability of using the Box-Cox transformed dependent variable. None of the coefficients of the air pollution variables is statistically significantly different from zero at the 0.05 level, and two of the coefficients have incorrect signs. Nonsignificance and incorrect signs of the coefficients of the air pollution variables are consistent features of the results of the econometric analysis carried out in this research. In the case of the hedonic price function, these features persist when additional principal components are added to the specification and when the pollution variables are transformed in various ways. Thus, the incorrect signs and nonsignificance of the coefficients of the pollution variables displayed in Table 4.1 are unlikely to be artifacts of the specification on which the tabulated results are based. Rather, there appears to be no statistically meaningful relation between house prices and air pollution in the Baltimore data set.

TABLE 4.1 -- PARAMETERS OF THE HEDONIC PRICE FUNCTION

Variable	Coefficient	Standard Error
Intercept	-1.23	0.673
Principal component no. 1	-0.167	0.0226
Principal component no. 2	0.926	0.0283
Principal component no. 3	-0.299	0.0202
Principal component no. 4	-0.0303	0.0207
Principal component no. 5	0.140	0.0165
Principal component no. 6	-0.0736	0.0355
Principal component no. 7	0.00256	0.0207
Principal component no. 8	0.108	0.0265
Principal component no. 9	0.101	0.0220
Principal component no. 10	0.0483	0.0190
Particulate matter	0.0106	0.00575
NO_2	-0.0106	0.00921
O_3	0.00118	0.00241
λ	0.506	0.0167
Variance of ε	0.252	0.0113

Prices are in tens of thousands of dollars. Pollutant concentrations are in $\mu g/m^3$.

As was discussed in Section 1.4, the causes of the poor performance of the pollution variables cannot be identified with certainty from the available data. Possible causes are errors in estimation of pollutant concentrations introduced by the interpolation procedure, the relatively low levels of pollutant concentrations in the Baltimore area during 1978, and lack of large variations in pollutant concentrations among locations in the Baltimore area. None of these possible causes can be tested convincingly using the available data. In particular, the magnitudes of the errors introduced by the interpolation procedure cannot be evaluated in the absence of concentration measurements in the vicinities of houses far from the fixed-site monitors.

Table 4.2 shows the estimated coefficients of the willingness-to-pay function obtained with the new and standard models as well as the values of the two models' log likelihood functions. The log likelihood of the new model exceeds that of the standard model by 1180.73. As in the discussion in Section 3.4, this large difference in log likelihood values implies strong rejection of the standard model in favor of the new one.[2]

The estimated value of the proportionality constant relating the hedonic price function to transactions costs, maintenance costs, and resale value is 1.074. This is consistent with the assumed rate of house appreciation of 8.5 percent per year (see Section 3.2), annual maintenance costs of one percent of a house's value, and transaction costs (including moving costs) of 5 percent of a house's value. Thus, the estimated value of the proportionality constant is reasonable.

It can be seen from Table 4.2 that there are large differences in the coefficient estimates obtained from the new and standard models. The effects of these differences on estimates of buyers' willingness to pay for changes in houses' attributes will be illustrated later in this section. The air

TABLE 4.2 -- PARAMETERS OF THE WILLINGNESS TO PAY FUNCTION

	New Model		Standard Model	
Variable	Coefficient	Standard Error	Coefficient	Standard Error
Intercept	14.6	9.92	6.53	1.72
Principal component no. 1	-0.0826	0.196	0.0743	0.0471
Principal component no. 2	-0.854	0.282	-0.142	0.0436
Principal component no. 3	-0.257	0.146	0.176	0.0423
Principal component no. 4	0.170	0.174	-0.142	0.0524
Principal component no. 5	0.178	0.181	0.155	0.0376
Principal component no. 6	0.116	0.156	0.0571	0.1184
Principal component no. 7	-0.586	0.165	0.0891	0.0568
Principal component no. 8	1.03	0.264	-0.0697	0.0588
Principal component no. 9	0.371	0.188	0.118	0.0517
Principal component no. 10	0.210	0.209	0.0434	0.0501
Particulate matter	0.109	0.0650	-0.0123	0.0174
NO_2	-0.140	0.123	-0.0061	0.0241
O_3	-0.0273	0.0252	-0.0131	0.0052
Log likelihood	-570.92		-1751.65	

Willingness to pay is in tens of thousands of dollars. Pollutant concentrations are in $\mu g/m^3$.

pollution coefficients in the new model are not statistically significantly different from zero at the 0.05 level, and the coefficient of particulate matter has the incorrect sign. The poor performance of the air pollution variables in the new model's willingness-to-pay function is consistent with their performance in the hedonic price function (i.e., it would be surprising to find that house buyers are willing to pay to achieve low values of the model's pollution variables but that these variables do not affect houses' sale prices). This consistency, as well as the poor performance of the pollution variables in a variety of specifications of the hedonic price function, suggests that the poor performance of these variables is attributable to the characteristics of the available data and is not an artifact of the specifications adopted here.

The air pollution coefficients of the willingness-to-pay function estimated with the standard model all have correct signs, and the O_3 coefficient is nominally statistically significantly different from zero. However, given the virtual certainty that the standard model is erroneous, little importance can be assigned to the signs, magnitudes, or nominal significance levels of the coefficient estimates obtained with this model. Indeed, as will be discussed presently, acceptance of the coefficient estimates obtained with the standard model could lead to highly misleading inferences.

The new and standard models are compared graphically in Figures 1.1 and 1.2. As was discussed in Section 1.4, these figures confirm the formal statistical finding that the new model provides a better explanation of the operation of the Baltimore area housing market than does the standard model. The figures give no indication that the new model is seriously misspecified, despite the simple specifications that have been used for the hedonic price and willingness-to-pay functions. As was also discussed in Section 1.4, the differences in fit illustrated in Figures 1.1 and 1.2 are reflected in the

ability of the two models to predict houses' sale prices. The new model has a root-mean-square (RMS) prediction error of $2,830, whereas the standard model has a RMS prediction error of $11,360.

The differences between the new and standard models' predictions of buyers' willingness to pay for changes in certain arbitrarily selected attributes of houses are illustrated in Table 1.1. These estimates were obtained by computing the changes in the values of the principal components, X, corresponding to changes in attribute values. The computation relies on the fact that the principal components are known functions of the attributes. As was discussed in Section 1.4, the differences between the estimates obtained from the two models are very large, and the estimates obtained from the standard model are highly unreasonable. In particular the estimates obtained from the standard model all have incorrect signs. Although it is impossible to determine from the available data whether the estimates obtained from the new model are approximately correct, they all have correct signs and reasonable magnitudes.

4.4 CONCLUSIONS

The standard hedonic model of consumers' choices in housing markets forms the basis of much environmental benefits analysis. However, it is inconsistent with the market mechanisms that casual observation suggests operate in real housing markets. The research reported here has provided strong econometric evidence that the standard model is, in fact, misspecified. An alternative model has been proposed. This model is consistent with the main characteristics of the market mechanisms that casual observation suggests operate in real housing markets. The econometric evidence indicates that the alternative (new) model provides a considerably better description of the operation of housing markets than does the standard model. The econometric evidence also indicates that the new and standard models give vastly different estimates of house buyers'

willingness to pay for housing amenities. Indeed, the willingness-to-pay estimates obtained in this research from the standard model are highly unreasonable, whereas those obtained from the new model are reasonable. In summary, the new and standard models are clearly distinguishable in terms of their underlying assumptions, empirical performance, and practical consequences. The performance of the new model is clearly superior to that of the standard model.

FOOTNOTES FOR CHAPTER 4

1. Mathematically,

$$W^* = W + \gamma E(p),$$

where W is the willingness-to-pay function, E(p) is the expected value of the hedonic price function, and γ is a constant parameter.

2. The significance test of equation (3.14) continues to indicate strong rejection of the standard model. In this case, application of the test requires making the approximation that sampling errors in the estimated value of the hedonic price function have a negligible effect on the test's outcome. Because the difference between the log likelihoods of the new and standard models is very large, this approximation is not likely to be severe.

REFERENCES

1. W. Alonso, "Location and Land Use," Harvard University Press, Cambridge, MA, 1953.

2. B. Bender, T.J. Gronberg, and H. Hwang, Choice of functional form and the demand for air quality, Review of Economics and Statistics, 62. 638-643, 1980.

3. K.E. Berndt, B.H. Hall, R.E.Hall, and J.A. Hausman, Estimation and inference in nonlinear structural models, Annals of Economic and Social Measurement, 3/4, 653-665, 1974.

4. D.S. Brookshire, M. Thayer, W. Schultze, and R.C. D'Arge, Valuing public goods: a comparison of survey and hedonic approaches, American Economic Review, 72, 165-177, 1982.

5. J.N. Brown and H.S. Rosen, On the estimation of structural hedonic price models, Econometrica, 50, 765-768, 1982.

6. H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, NJ, 1946.

7. R.A. Dubin and A.C. Goodman, Valuation of crime and education neighborhood characteristics in a hedonic housing price model, Population and Environment, 5, 66-81, 1982.

8. B. Ellickson, An alternative test of the hedonic theory of housing markets, Journal of Urban Economics, 9, 56-79, 1981.

9. J. Galambos, "The Asymptotic Theory of Extreme Order Statistics," John Wiley and Sons, New York, 1978.

10. E.J. Gumbel, "Statistics of Extremes," Columbia University Press, New York, 1958.

11. D. Harrison , Jr., and D.L. Rubinfeld, Hedonic housing prices and the demand for clean air, Journal of Environmental Economics and Management, 5, 357-369, 1978.

12. J.L. Horowitz, Statistical comparison of non-nested probabilistic discrete choice models, Transportation Science, 17, 319-350, 1983.

13. Y.M. Ioannides, Market allocation through search: equilibrium adjustment and price dispersion, Journal of Economic Theory, 11, 247-262, 1975.

14. S.R. Lerman and C.R. Kern, Hedonic theory, bid rents, and willingness to pay: some extensions of Ellickson's results, Journal of Urban Economics, 13, 358-363, 1983.

15. M. Li and H.J. Brown, Micro-neighborhood externalities and hedonic housing prices, Land Economics, 57, 125-141, 1982.

16. R.F. Muth, "Cities and Housing," The University of Chicago Press, Chicago, 1969.

17. J.P. Nelson, Residential choice, hedonic prices, and the demand for urban air quality, Journal of Urban Economics, 5, 357-369, 1978.

18. S. Rosen, Hedonic prices and implicit markets: product differentiation in pure competition, Journal of Political Economy, 82, 34-55, 1974.

19. W.C. Wheaton, A bid rent approach to housing demand, Journal of Urban Economics, 4, 200-217, 1977.

APPENDIX A

PROOF THAT PR(H SOLD) $\leq$ 1 IN (2.28)

To establish that Pr(h sold) < 1 in (2.28), it is necessary to specify the process through which sellers decide to accept or reject offers in sufficient detail to insure that a seller accepts at most one offer for a house. It is assumed here that if the seller of house H receives a single offer of size p for that house, he accepts this offer with probability $Q_h^*(p)$, where $Q_h^*(P_{ah}) = 1$. It is assumed that if several offers are received simultaneously and the highest offer is p, this offer is accepted with probability $Q_h^*(p)$ and the others are rejected. As explained in footnote 4 of Chapter 2, the probability that two or more offers of size P_{ah} are received simultaneously is assumed to be negligible.

Suppose that household i makes an offer of $b < P_{ah}$ for house h. The probability that this is the highest offer made is $\prod_{j \neq i} \int_0^b g_{ojh}(p)\,dp$, where g_{ojh} is given by (2.22) and the product extends over all households $j \neq i$ in the housing market. The probability $Q_h(b; i)$ that household i's bid is accepted is then

$$Q_h(b; i) = Q_h^*(b) \prod_{j \neq i} \int_{-\infty}^{b} g_{ojh}(p)dp. \tag{A1}$$

Since the probability density that i makes an offer of b is $g_{oih}(b)$, the probability density that house h is sold to household i at price $p_s < P_{ah}$ is

$$g_h(i, p_s) = Q_h(p_s; i)g_{oih}(p_s)$$

$$= Q_h^*(p_s)g_{oih}(p_s) \prod_{j \neq i} \int_{-\infty}^{p_s} g_{ojh}(p)dp. \tag{A2}$$

Equation (A2) is analogous to equation (2.25) of the text but with $Q_h(p_s)$ replaced by $Q_h(p_s;i)$.

The probability that h is sold at a price less than P_{ah}, regardless of buyer, is

$$\sum_i \int_{-\infty}^{P_{ah}} g_h(i,p)dp = \int_{-\infty}^{P_{ah}} Q_h^*(p)[(d/dp)\prod_i \int_{-\infty}^{p} g_{oih}(p')dp']dp$$

$$\leq \prod_i \int_{-\infty}^{P_{ah}} g_{oih}(p')dp' \quad , \tag{A4}$$

where the inequality follows from $Q_h^*(P) \leq 1$. But

$$\int_{-\infty}^{P_{ah}} g_{oih}(p')dp' \leq 1 - G_{oih}(P_{ah}) \ , \tag{A5}$$

where $G_{oih}(P_{ah})$ is given by (2.24). Therefore

$$\sum_i \int_{-\infty}^{P_{ah}} g_h(i,p)dp \leq \prod_i [1 - G_{oih}(P_{ah})]. \tag{A6}$$

Since the probability that the seller of h receives simultaneously two or more offers equal to P_{ah} is assumed to be negligible and since $G_{oih}(P_{ah})$ is the probability that household i makes an offer of P_{ah}, it follows that

$$\prod_i [1 - G_{oih}(P_{ah})] \simeq 1 - \sum_i G_{oih}(P_{ah}). \tag{A7}$$

Hence

$$\sum_i \int_{-\infty}^{P_{ah}} g_h(i,p)dp \leq 1 - \sum_i G_{oih}(P_{ah}). \tag{A8}$$

It follows from (2.24) and (2.26) that house h is sold to household i at price P_{ah} with probability $G_{oih}(P_{ah})$. Therefore, the probability that house h is sold at price P_{ah} regardless of the buyer is $\sum_i G_{oih}(P_{ah})$. Combining this result with (A8) yields

$$\Pr(\text{h sold}) = \sum_i \int_{-\infty}^{P_{ah}} g_h(i,p)dp + \sum_i G_{oih}(P_{ah}) \leq 1. \tag{A9}$$

APPENDIX B

CONTENTS OF THE DATA SET

The data elements available for each house are as follows:

Data Element	Data Source
Sale price	1
Asking price	1
Annual property tax	1
Attributes of house:	
Location (city or county)	1
Type of dwelling (attached or detached)	1
Number of rooms	1
Number of bathrooms	1
Basement (yes or no)	1
Swimming pool (yes or no)	1
Patio (yes or no)	1
Fireplace (yes or no)	1
Air conditioning (yes or no)	1
Garage (yes or no)	1
Age	1
Lot size	2
Floor space	2
Harmonic mean distance to major shopping locations	3
Harmonic mean distance to freeway on-ramps	3

Data Element	Data Source
Attributes of census tract in which house is located:	
Mean household income (1970 and 1980)	4
Elementary and high school students per household (1970 and 1980)	4
Percentage of population that is black (1970 and 1980)	4
Percentage of population over age 25 who have completed high school (1970 and 1980)	4
Median age of population (1970 and 1980)	4
Percentage of employed residents with blue collar occupations (1970 and 1980)	4
Percentage of housing units that are owner occupied (1970 and 1980)	4
Other neighborhood attributes:	
Six indices of school quality for the school district in which the house is located (1978)	5
Three indices of crime levels for the crime reporting district in which the house is located (1978)	5

Composite attributes:	
(Lot size) x (Elementary and high school students per household in 1980)	6
(Floor space) x (Persons per dwelling unit in 1980)	6
(Each of the six education indices) x (Elementary and high school students per household in 1980)	6
Air pollution variables:	
Annual average particulate matter concentration	7
Annual average NO_2 concentration	7
Annual maximum hourly O_3 concentration	7

Data Sources:

1. Central Maryland Multiple Listing Service
2. Baltimore City and County tax records.
3. Distances are airline distances.
4. U.S. Census fifth count for 1970 and standard tape file STF3A for 1980.
5. Dubin and Goodman [7].
6. The composite variables were created by combining house and neighborhood variables obtained from the sources listed above.
7. Interpolated from monitoring data obtained from the Maryland Bureau of Air Quality.

APPENDIX C

QUALITY CONTROL EVALUATION REPORT

The data used in this research were obtained from the records of the Central Maryland Multiple Listing Service, the governments of Baltimore City and County, the U.S. Census (fifth count for 1970 and standard tape file STF3A for 1980), and the Maryland Bureau of Air Quality. No original data acquistion or production were undertaken, and no original measurements were made. The data obtained from the above listed sources were edited to eliminate coding errors and records with missing data elements. Further checks of the accuracy of the data were not possible since the data pertain to past years and sources of similar data for those years do not exist.